DOCUMENT RESUME

ED 201 932 CG 015 159

AUTHOR Priest, Robert F.; Adams, Jerome
TITLE Validity, Not Bias, in Teacher Ratings of Leadership.
PUB DATE Sep 80
NOTE 7p.; Paper presented at the Annual Convention of the American Psychological Association (88th, Montreal, Quebec, Canada, September 1-5, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Admission Criteria; *College Applicants; *Competitive Selection; High Schools; *Leadership Qualities; Leadership Training; Military Schools; Personality Traits; *Predictive Validity; *Student Evaluation; Student Leadership; *Teacher Attitudes
IDENTIFIERS Military Academy (West Point) NY

ABSTRACT

Although high school teacher evaluations of student non-cognitive attributes, such as leadership ability, are frequently considered in selecting applicants for competitive colleges, little empirical research exists on the reliability, validity, and bias of teacher ratings. The United States Military Academy at West Point requires that a set of school officials rate applicants on 10 non-cognitive items designed to measure potential for leadership in the highly structured Academy environment. Measures of leadership potential and student self-reports of athletic participation and extracurricular activities are also considered. Leadership performance at the Academy is assessed by cadet officers, Army officers, and upperclassmen. For a sample of 15,557 applicants to West Point, teacher ratings of leadership potential were related to measures of extracurricular participation and academic performance. Such ratings were found to be unbiased with respect to race and sex. Measures of leadership performance in the first two years at West Point correlated significantly with teacher ratings. Other alternative measures were not as valid as the teacher ratings. (NRB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201932

Validity, Not Bias, in Teacher Ratings of Leadership*

Robert F. Priest, Ph.D.
MAJ Jerome Adams, Ph.D.

U.S. Military Academy

Abstract

High school teacher evaluations of non-cognitive attributes, such as leadership ability, are frequently used in selecting students for competitive colleges. This study shows that, in a sample of 15,557 applicants to West Point, teacher ratings of leadership potential are related in meaningful ways to measures of extracurricular participation and academic performance. Such ratings are unbiased with respect to race and sex. Measures of leadership performance in the first two years at West Point correlate significantly with teacher ratings in high school. Though validities are low, utility may be high. Furthermore, no alternative measure works as well.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert F. Priest

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CG 015159

*This paper is prepared for the Division of Industrial Organizational Psychology of the American Psychological Association for the 1980 annual meeting in Montreal, 1-5 September. It represents the opinions of the authors only, not the official position of the U.S. Military Academy, or the U.S. Army.

Problem

High school teacher evaluations of student attributes, both cognitive and non-cognitive, are frequently considered in selecting among applicants at competitive colleges (Aleamoni, 1972; Greenberg & O'Brien, 1976). Texts on selection sometimes imply that recommendations, no matter how obtained, are biased against minorities or women, or are totally useless as predictors (Guion, 1965, p14; Stone & Kendall, 1956, p]75). However, except for data which documents the use of teacher evaluations in admissions to medical school (Rainer & Luecking, 1974), law school (Pipkin & Katsh, 1976) or to graduate school (Lewis, 1972), there has been no empirical research on the reliability, validity, or unbiasedness of teacher ratings.

There is no question that standardized tests are superior to teacher ratings as measures of cognitive abilities needed in college (Cleary et al., 1975; Stanley, 1976). In measuring an important non-cognitive ability, such as leadership, no satisfactory standardized test has been developed. While many colleges develop leadership, few institutions make the selection and development of leadership ability a goal as explicitly as the service academies. Given the importance of leadership to these academies, they have taken care to develop systematic methods for assessing the leadership ability of high school students. The present research examines the reliability, validity, and bias of high school teacher ratings predicting leadership performance at the U.S. Military Academy.

Method

Teacher Ratings

The Academy requires ratings from a set of school officials who are likely to have diverse perspectives on the performance of a student: an English teacher, a math teacher, a guidance counselor and a coach or PE teacher. They rate the student on 10 non-cognitive items designed to measure potential for leadership in the highly structured academy environment. Each item is a phrase such as "ability to work toward group goals in a subordinate position", or "ability to gain respect of peers", rather than the one-word trait list used by some other colleges for counselor recommendations. Each item is rated on a five point scale. The resultant Faculty Appraisal Score (FAS) has a range of 40 to 740. Its estimated reliability is .479 (coefficient alpha).

Other Measures of Leadership

The Academy considers two other measures of leadership potential in the high school student, both based on the self reported extracurricular participation of the student. A measure of athletic participation (AT), with scores ranging from 200 to 800, assigns higher scores to students who participate as team captains and letter in different sports, lower scores to non participants or non-letter holders. A measure of non-athletic extracurricular activities (EX), gives highest scores to class presidents, moderate scores to club presidents, low scores to nonparticipants. Both the AT and EX scores are machine scorable from the student's report of his extracurricular performance. An estimate of internal consistency is not appropriate because of configural scoring, but test-retest reliability of the form would be expected to be extremely high. The Academy combines all three measures of high school student leadership (FAS, EX, and AT) into an equal-weighted composite with a composite reliability of .319. This is known as the Leadership Potential Score (LPS).

Leadership Performance

Leadership performance at the Academy is assessed by a composite of assessments by cadet officers in the Chain of Command (COC), by army officers who act as Company Tactical Officers (TAC), and by cadets in the company who are in a class above the individual being rated. Cadets are rated every semester. In this analysis, performance ratings covering the first two years of the Class of 1981 were used. The four TAC ratings (one per semester) have a composite reliability of .559; the four COC ratings have a composite raliability of .764; the reliability of the ratings from all sources (STATOT) over the two years is .915. A more complete description of the measures used in this study is in Priest (1980).

Results

Investigating Bias in the Applicant Population

It was hypothesized that the high school faculty rater would be influenced by several variables: by the academic performance of the student as measured by his rank in class (HSR); and by the student's extracurricular performance (AT, EX). Furthermore, because students from large high schools applying to West Point usually have stiffer competition from peers than students from small high schools, and also because prior research suggested that Class Size (SIZE) affected teacher ratings (Orvick, 1973), it was predicted that teacher ratings also would be affected by class size. Sex and minority group membership were also included as predictors of FAS.

An analysis of the predictors of FAS was made, in three successive groups of applicants to West Point. The results are presented in Table 1. In all cases, the hypotheses above were supported at the .001 level. The correlations are, at best, modest in size. After correcting for attenuation due to unreliability in the FAS, the corrected validity coeficients show that most of the variance of teacher ratings cannot be explained by academic performance, extracurricular participation, or class size. This supports the concept that the FAS provides information on student leadership which is not merely a reflection of objective information obtainable from student records.

The multiple regression analysis shown in Table 1 was extended to include both sex and minority as predictors of FAS. With HSR, AT, EX and SIZE held constant, neither sex nor minority group membership added significantly to the prediction of FAS in any of the three years. Given the large sample size, which allows for the detection of even small amounts of gender or race bias, these negative results are especially meaningful. One can confidently conclude that the FAS and the high school officials who use it are not biased toward one race or one gender in evaluating the leadership potential of candidates for a Military Academy appointment. This is an encouraging result given the strong demand from government and interest groups concerned with potential bias in selection criteria.

Table 1

Corrlates of
Faculty Appraisal Score in
the Applicant Population

Predictor	Class Year	Raw Validity	Corrected[a] Validity
HSR	81	.275	.397
	82	.266	.384
	83	.235	.339
AT	81	.246	.355
	82	.233	.337
	83	.204	.295
EX	81	.200	.289
	82	.166	.240
	83	.154	.222
SIZE	81	.057	.082
	82	.118	.170
	83	.099	.143
Multiple R (4 steps)	81	.415	.600[b]
	82	.405	.585[b]
	83	.355	.512[b]

N's for CY 81, 82 and 83 are 5808, 5116 and 4633 respectively.

[a] Corrected for alteration due to unreliability in FAS.

[b] Since the Multiple R is a product moment correlation between FAS and a weighted combination of 4 variables, the correction for attenuation can be applied.

Predictive Validities

Table 2 shows the validity of four measures of high school leadership in predicting the composite leadership performance of cadets over a two-year interval at West Point. The coeficients have been corrected for attenuation due to unreliability. The results show that teacher ratings (FAS) are more valid as predictors of leadership than the two measures of extracurricular participation. The composite Leadership Potential Score (LPS) is little better than the FAS alone. The disattenuated validity coeficients show the results to be expected if both predictors and criteria were perfectly reliable measures: Thus, even at best, high school leadership shows little variance in common with leadership at the Academy.

Table 2

Validities After Correcting for Attenuation
in Predictor and Criterion

Predictor	Criterion TAC	COC	STATOT
AT	.08	.14*	.13*
EX	.12	.14*	.13*
FAS	.35*	.34*	.36*
LPS	.35*	.43*	.41*

*When the original (uncorrected) correlation was tested for significance, it was significant at the .001 level. All r's based on at least 1084 cases in the Class of 1981.

Discussion

The results are consistent with Cleary, Humphries, Kendrick and Wesman's (1975) assessment of research on measurement of non cognitive measures in general. It is important to do research on them, but optimism is not warranted. Teacher leadership ratings are not as reliable as mental test scores, and while they have some relation to performance of leaders in a new setting, most of the reliable variance of college level leadership cannot be explained by reliable variance in leadership assessed by high school teachers. One should note however, that while teacher ratings are not very good in an absolute sense, they are the best measure we have, clearly better than two quantified measures of extracurricular participation.

The results are consistent with the work of those social psychologists who emphasize a situational theory of leadership. There is no a priori reason why requirements for leadership in the average American high school should have an extremely close correspondence to the requirements for leadership in the highly structured environment of a Military Academy.

Given the importance of leadership to the Academy, however, there is a justification for continuing teacher ratings as a measure of leadership, Even though validities are low. As Schmidt, Hunter, McKenzie and Mulrow (1979) have recently demonstrated, the utility of a test is not solely dependent on its validity. To demonstrate the utility of teacher ratings would require estimates of the dollar value of high and low ranking cadet leaders to the Army. Such a demonstration obviously requires further research.

References

Aleamoni, L. The vagaries of Selection - 1. Selection, Placement and Proficiency in American Colleges and Universities. The National ACAC Journal 1972, 17, 7-12.

Cleary, T.A., Humphries, L. Kendrick, S. & Wesman, A. Education uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-40.

Greenberg, A., & O'Brien, A. College Admissions 1975. Paper presented at the annual meeting of the American Personnel and Guidance Association (Chicago, 1976). Abstracted in Eric, ED128671.

Guion, R. Personnel testing. New York: McGraw Hill, 1965.

Lewis, L. On the Genesis of Gray-Flanneled Puritans. AAUP Bulletin 1972, 58, 21-29.

Orvick, J. The predictive validity of premilitary performance ratings by high school personnel, Journal of Applied Psychology, 1973, 57, 88-89.

Pipkin, R., & Katsh, E. Undergraduate legal studies and law school gatekeepers. Journal of Legal Education, 1976, 28, 103-111.

Priest, R. Reliability, Validity and Bias in Measures of High School Leadership. Office of the Director of Institutional Research, West Point, NY, Report 80-006, May 1980.

Rainey, M., & Luecking, D. Admissions Committee Gatekeeping: The case of the female applicant to medical school. Paper read at the Midwest Sociological Society, Omaha, April 1974. Abstracted in College Student Personnel Abstracts, Spring 1975, 10, 284-285.

Schmidt, F., Hunter, J.E., McKenzie, R., & Muldrow, T. Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 1979, 64, 609-627.

Stanley, J. Test better finder of great math talent than teachers are. American Psychologist, 31, 313-314.

Stone, C.H., & Kendall, W.E. Effective personnel selection procedures. Englewood Cliffs, N.Y.: Prentice-Hall, 1956.